EEU44C04 / CS4031 / CS7NS3 / EEP55C27
Next Generation Networks

Queuing systems

Nicola Marchetti
nicola.marchetti@tcd.ie

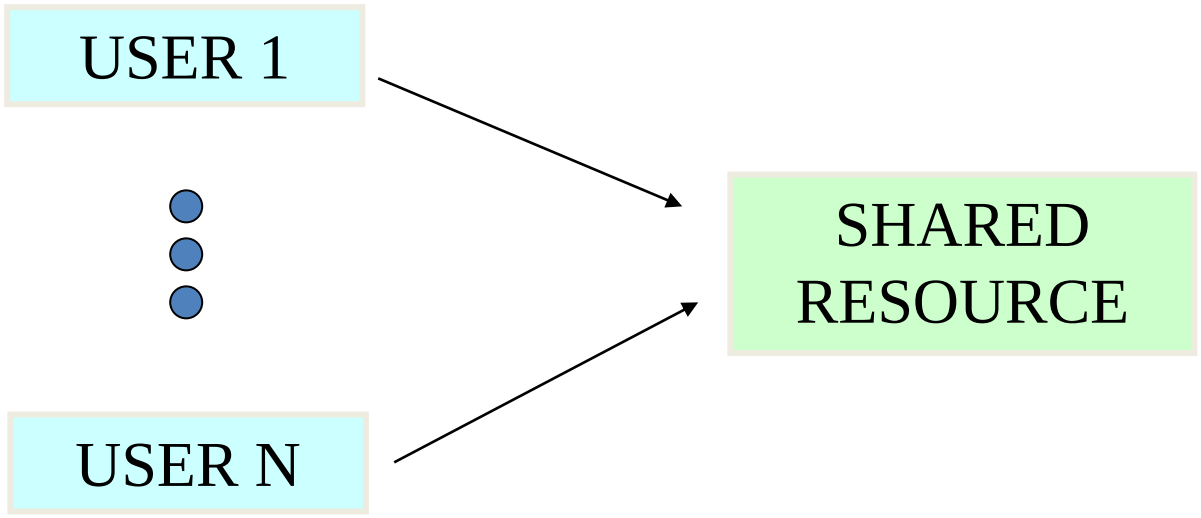
What is a Queue?

From Merriam-Webster's Collegiate Dictionary:

Main Entry: [2]queue
Function: *verb*
Date: 1777
transitive sense: to arrange or form in a queue
intransitive sense: to line up or wait in a queue -- often used with *up*

Motivation

- Analytical models based on queuing theory can often be used to predict the effects of some change in load or design



Examples (1)

(a) Time-shared computers

- Programs ⇔ CPU, Disk, I/O

(b) Statistical Multiplexer / Concentrator

- Packet-based
  - Packets ⇔ Link
- Channel-based
  - Calls ⇔ Channels

Examples (2)

(c) Multiple Access Network

- Ethernet LAN
  - Frames ⇔ Medium (Coaxial, Fiber)
- Wireless Network
  - Frames ⇔ Wireless Medium

(d) Web Access

- Clients ⇔ Server

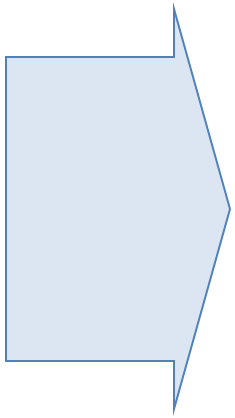
What does queuing theory study?

OF USERS

ARRIVAL CHARACTERISTICS

SERVICE CHARACTERISTICS

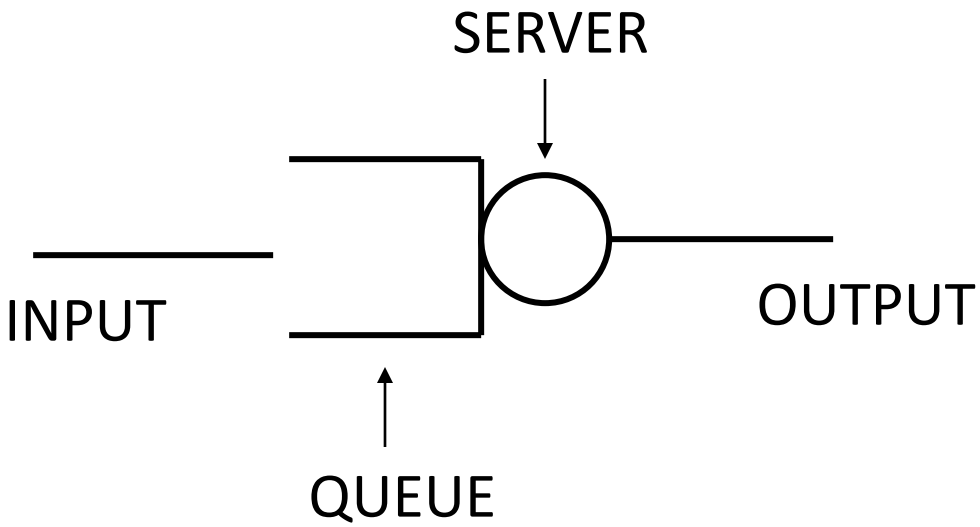
RESOURCES



PERFORMANCE

- waiting time
- blocking

Elements of a queuing system



Model (1)

- Customers from some population arrive at the system at random *arrival times*
- λ is the customer *arrival rate*
- Queuing system has c identical *servers*
- The j^{th} customer seeks a service that will require s_j units of *service time* from one server
- If all servers are busy, arriving customer joins a queue until a server is available

Model (2)

- *Service discipline* specifies the order in which customers are selected from the queue
  - ex: FIFO, LIFO, priority, fair queuing, …

- *Waiting time* t_{Qj} is the time j^{th} customer is made to wait between entering the system and entering service

- *Total delay in the system* $\tau_j = t_{Qj} + s_j$

- n ≡ number of customers in the system (a r.v.)
- n_q ≡ number of customers in the queue (a r.v.)

a/b/m/K notation (Kendall's notation)

- a = type of arrival process
  - M (Markov) denotes Poisson arrivals, so interarrival times are iid, exponential random variables
- b = service time distribution
  - M (Markov) denotes exponentially-distributed
  - D (Deterministic) denotes constant service times
  - G (General) denotes iid service times following some general distribution
- m = number of servers
- K = maximum # of customers allowed in the system

Problem

Identify the queuing model for the following system:

- Inter-arrival times are independent and identically distributed, exponential RV
- Service times are constant
- More than one server is available
- The first being served is the last that arrived

M / D / m with **LIFO** service discipline